MANUAL DE USUARIO

Sistema BDGN Lucene

CRISTIAN RÍOS VILLADA

JOHN JAIRO SILVA ZULUAGA

Medellín

Universidad Eafit

2008

TABLA DE CONTENIDO

INTRODUCCIÓN *1*

REQUERIMIENTOS PARA EJECUCIÓN *2*

HARDWARE **2**

SOFTWARE **2**

PASOS PARA EJECUTAR LA APLICACIÓN *3*

MODULOS FUNCIONALES *4*

MODO DE BÚSQUEDA **4**

Lenguaje de Búsqueda 5

Operadores 5

Modificadores de texto 6

Búsqueda básica 7

Búsqueda avanzada 8

Ver resultados 9

Ver detalle 10

MODO ADMINISTRACIÓN **11**

MENU PRINCIPAL 12

ADMINISTRAR REPOSITORIOS 13

Seleccionar un repositorio 14

Crear repositorio 15

Eliminar repositorio 15

INDEXACIÓN 16

Crear un documento 16

Indexar documentos 19

Indexar documentos XML Dublin Core 20

BUSQUEDA 21

Búsqueda básica 21

Búsqueda avanzada 22

Ver resultados 23

Ver detalle 24

Eliminar documento 25

INTRODUCCIÓN

BDNG Lucene es una aplicación web basada en el núcleo BDNG (Biblioteca Digital Nueva Generación) de la universidad EAFIT con el fin de crear una librería virtual donde los usuarios puedan ingresar archivos con sus respectivos metadatos y las opciones de Buscar, Actualizar y Eliminar dichos documentos.

Los procesos de almacenamiento e indexación de la información son realizados con la ayuda de la librería de código abierto Lucene, la cual sirve de herramienta para convertir la aplicación en un potente y eficaz motor de búsqueda.

Este documento se encarga de describir funcionalmente el sistema asi como el proceso de instalación. Los detalles de diseño y estructura se encuentran en el manual técnico.

REQUERIMIENTOS PARA EJECUCIÓN

HARDWARE

Mínimo:

- **Procesador:** 700 MHz
- **RAM:** 512 MB

Recomendado:

- **Procesador:** 1.8 GHz
- **RAM:** 1 GB

SOFTWARE

- Apache Tomcat 6 (http://tomcat.apache.org/download-60.cgi)
- Java 6. (http://java.sun.com/javase/6/)
- **Browser:** Probado en Internet Explorer, Mozilla Firefox o Google Chorme.

PASOS PARA EJECUTAR LA APLICACIÓN

1. Descargar e instalar el Apache Tomcat 6.
2. En el CD del proyecto BDNG Lucene, se encuentra el archivo BDNGlucene.war
3. Copiar dicho archivo en la carpeta *tomcat*/webapps la cual se encuentra en la ruta donde fue instalado el tomcat.

   O abrir el Tomcat Manager (http://127.0.0.1:8080/manager/html) y cargar el dicho archivo en la sección Desplegar - Archivo WAR a desplegar. Seleccionar la ruta del archivo y oprimir el botón *desplegar*.

   En ambos casos el proyecto debe aparecer en la sección de Aplicaciones del Tomcat Manager.
4. Luego se debe abrir el browser preferido por el usuario y digitar alguna de las siguientes rutas:

   - http:// localhost:8080/luceneweb/
   - http://127.0.0.1:8080/luceneweb/

   Ambas rutas corresponden a la configuración por defecto del servidor local de Tomcat, si dicha configuración, tanto la ruta como el puerto fueron cambiados, entonces el usuario deberá digitar dicha ruta y el nombre del proyecto */BDNGlucene*

MODULOS FUNCIONALES

El sistema posee 2 interfaces, una de búsqueda y otra de administración. La primera es un modo de solo consulta por medio de búsqueda básica y búsqueda avanzada; en la segunda el usuario puede agregar documentos, indexar directorios completos, indexar directorios de documentos xml Dublin Core y administrar los repositorios.

Por defecto, al digitar la url de ingreso (mencionada en paso 4 de la sección anterior), el sistema despliega inmediatamente el modo de búsqueda.

MODO DE BÚSQUEDA

Se puede realizar una búsqueda básica directamente o se puede ingresar a realizar una búsqueda avanzada.

Pero antes de entrar en detalle es conveniente realizar una especificación del lenguaje de búsqueda que soporta el API Lucene.

Lenguaje de Búsqueda

Operadores.

Dentro de los 2 tipos de búsqueda, el usuario puede ingresar términos y operadores lógicos en sus campos de búsqueda:

Un término individual es una sola palabra, como “test” o “hello”.

Una frase es un grupo de palabras rodeado por comillas, como “hello dolly”.

Puede buscar en cualquier campo escribiendo el nombre del campo seguido de dos puntos ":" y, después, el término que está buscando.

Como ejemplo, supongamos que un índice Lucene contiene dos campos, título y texto, y que texto es el campo predeterminado. Si quiere encontrar el documento titulado “The Right Way”, que contiene el texto “don't go this way”, puede escribir:

- title:"The Right Way" AND text:go
- title:"Do it right" AND right. Como el texto es el campo predeterminado, el indicador de campo no es necesario.

Sin embargo, para evitarle este tipo de complicaciones al usuario, la busqueda avanzada implícitamente se encarga de construir dicho query, y solo basta con seleccionar los campos en los que se debe buscar y los criterios de búsqueda.

Nota: El campo sólo es válido para el término al que precede directamente, de forma que

- title:Do it right

Sólo encontrará "Do" en el campo del título. Encontrará "it" y "right" en el campo predeterminado (en este caso, el campo de texto).

Modificadores de texto

Lucene es compatible con la modificación de términos de consulta para proporcionar un amplio rango de opciones de búsqueda.

- Búsquedas comodín

Lucene es compatible con búsquedas de caracteres comodín individuales y múltiples.

Para realizar una búsqueda comodín de un solo carácter, use el símbolo "?".

Para realizar una búsqueda comodín de varios caracteres, use el símbolo "*".

La búsqueda comodín de un solo carácter busca términos que coincidan con el carácter individual que se ha sustituido. Por ejemplo, para buscar "text" o "test" puede usar la búsqueda: te?t

La búsqueda comodín de varios caracteres busca 0 ó más caracteres. Por ejemplo, para buscar *test*, *tests* o *tester*, puede usar la búsqueda: test*

También puede usar las búsquedas comodín en medio de un término. te*t

Nota: No puede usar un símbolo * ni ? como primer carácter de una búsqueda.

- Búsquedas difusas

Lucene es compatible con búsquedas difusas basadas en el algoritmo de la Distancia Levenshtein o en el algoritmo de Distancia de edición. Para realizar una búsqueda difusa, use el símbolo de la tilde, "~", al final de una

búsqueda de una sola palabra. Por ejemplo, para buscar un término que se deletree de forma parecida a "roam" use la búsqueda difusa: roam~

Esta búsqueda encontrará términos como “foam” y “roams”.

- Operadores booleanos

Los operadores booleanos permiten que los términos se combinen mediante operadores lógicos. Lucene acepta AND, "+", OR, NOT y "-" como operadores booleanos (Nota: los operadores booleanos deben estar COMPLETAMENTE EN MAYÚSCULAS).

Para mas información acerca del tipo de búsquedas que se pueden realizar por medio de Lucene leer el siguiente documento:

- Sintaxis del Analizador de búsquedas de Lucene:

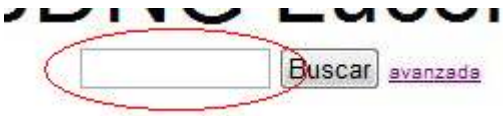
  http://www.ehu.es/p200-content/es/contenidos/informacion/ayuda_busquedas/es_def/adjuntos/Lucene-Overview.pdf

  (De la página de la Universidad del Pais Vasco, pero no tiene autor ni año de publicación, sin embargo es un material didáctico que comprende las búsquedas que se pueden realizar por medio de Lucene).

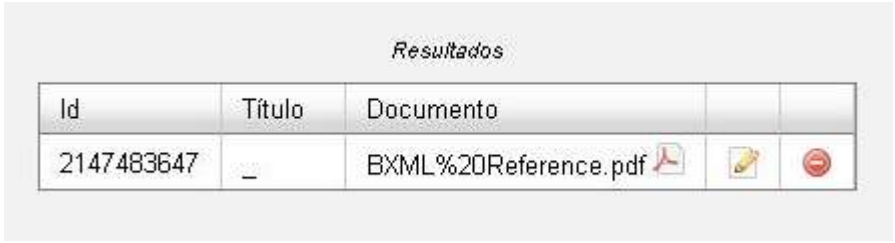
Búsqueda básica

Esta búsqueda consiste en ingresar los parámetros de búsqueda y el sistema se encargará de retornar los documentos coincidentes sin importar en que metadato o sección del documento se encuentran las palabras a buscar.

Para esta búsqueda se debe ingresar la(s) palabra(s) a buscar y oprimir el botón buscar.



El sistema retornará los resultados que coincidan con esta búsqueda y los campos de Id, Titulo y Documento.

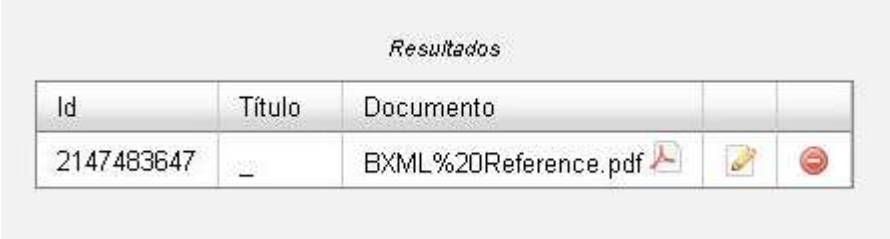


Búsqueda avanzada

Para ingresar a esta búsqueda se debe ingresar seleccionar el link de búsqueda avanzada, al lado del botón Buscar.

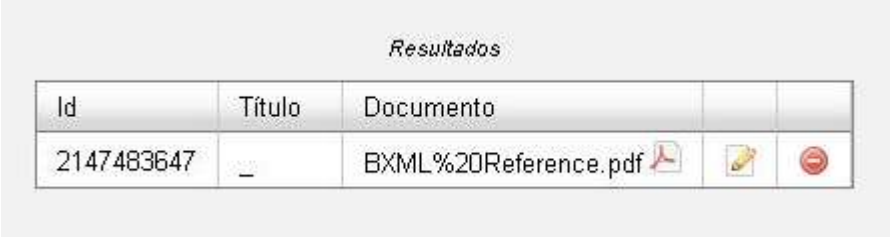
Para esta búsqueda se pueden seleccionar los campos específicos y las palabras que se desean buscar, y los operadores lógicos entre dichos campos para hacer más precisa la búsqueda.

Al oprimir el botón buscar se muestran los resultados según los criterios de búsqueda.

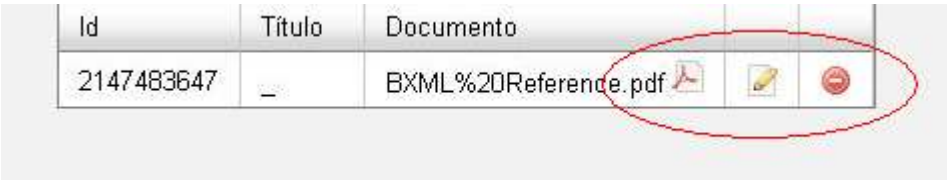


Ver resultados

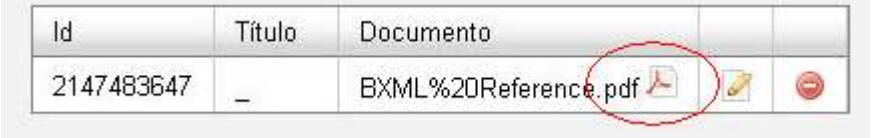
Sin importar el tipo de búsqueda, los resultados solo mostrarán algunos de los datos básicos asociados al documento.



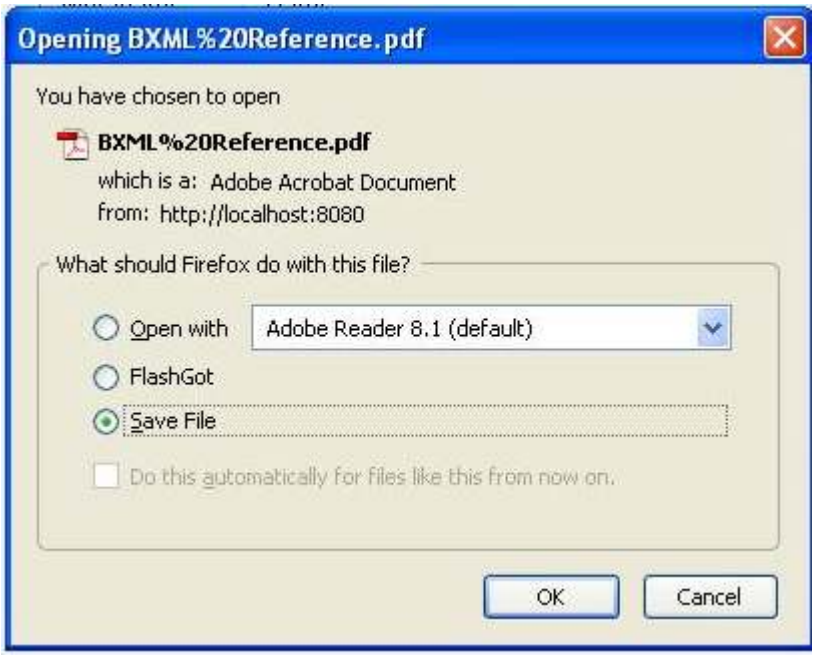
Cada registro encontrado tendrá la posibilidad de ver el detalle y descargar documento.



Para descargar el documento se debe dar click sobre el icono del documento.

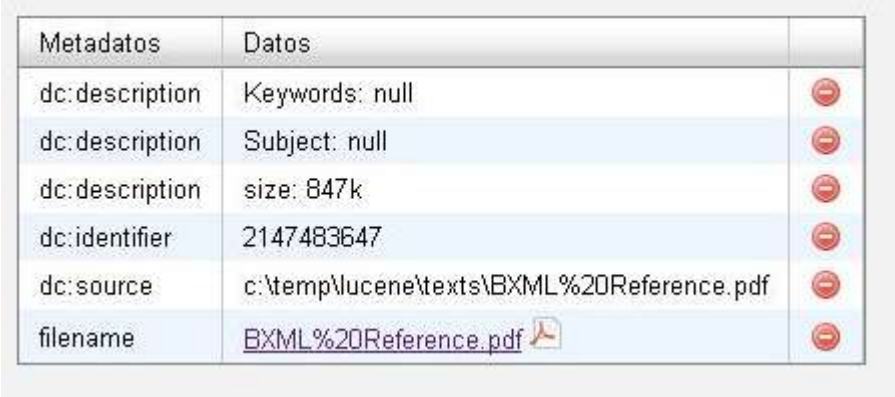


El sistema solicitará al usuario donde desea descargar el documento en la estación local, y allí será almacenado.



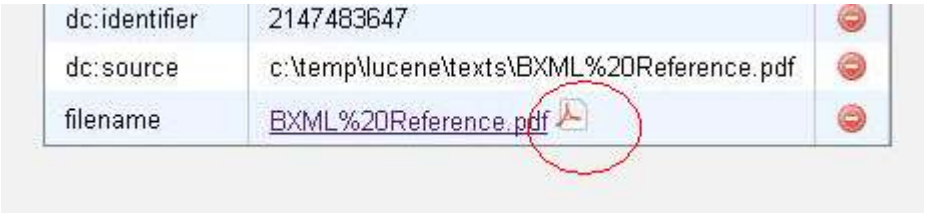
Ver detalle

En esta sección se muestran todos los datos que tiene el documento, si un dato se encuentra vacio entonces no se mostrará en el detalle.



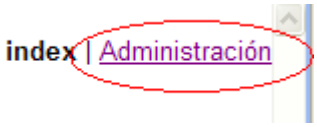
Metadatos	Datos
dc:description	Keywords: null
dc:description	Subject: null
dc:description	size: 847k
dc:identifier	2147483647
dc:source	c:\temp\lucene\texts\BXML%20Reference.pdf
filename	BXML%20Reference.pdf

Para descargar el documento se debe dar click sobre el icono del documento.



MODO ADMINISTRACIÓN

Para ingresar a este modulo se debe dar click en el link Administración ubicado en la parte superior derecha del modo de Búsqueda.



Para volver al modo de Búsqueda se debe seleccionar la opción Salir en la parte superior derecha del modo de Administración

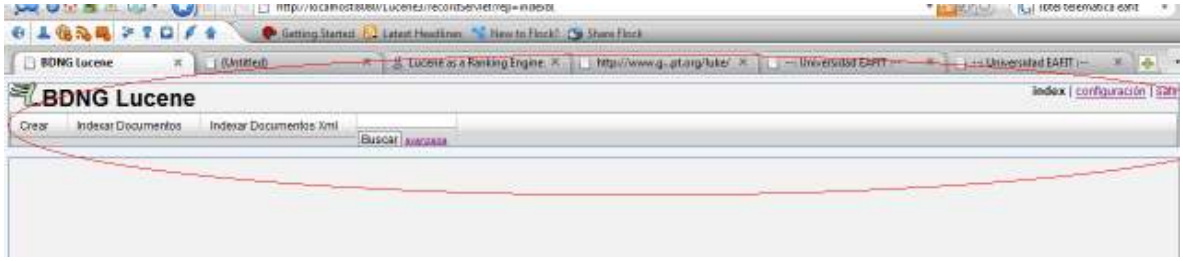
Cuando se ingresa a este modulo el sistema solicita seleccionar el repositorio en el cual se desea ejecutar la operación. También permite la opción de crear un repositorio nuevo.

El repositorio por defecto, en caso de que no se seleccione ninguno, aparecerá en la parte superior derecha. (Ver sección Administrar Repositorios para más detalles).

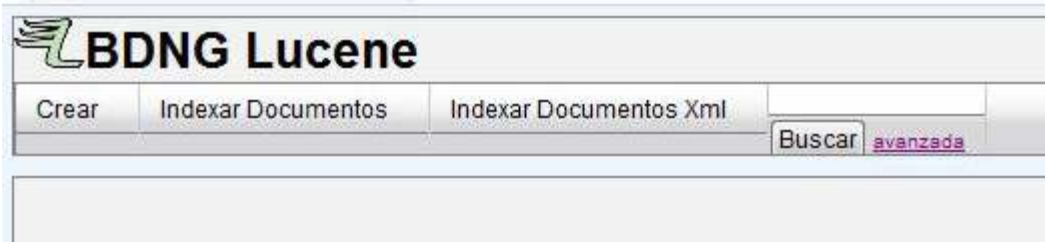
Nota: El repositorio por defecto se encuentra configurado en el archivo *configuration.properties* del sistema en la propiedad *index.* Para esto se debe tener acceso al código fuente del sistema. Para más detalles, ver el documento Manual Técnico.

MENU PRINCIPAL

El menú principal se encuentra en la parte superior de la página, parecida a una barra de herramientas.



En la parte superior izquierda se encuentran las operaciones a realizar sobre el índice seleccionado. (El índice seleccionado aparece en la parte superior derecha). Tanto las opciones para indexar como para buscar se encuentran en la misma sección.



En la parte superior derecha aparece el índice seleccionado, la opción de configuración (Administrar repositorios) y la opción de salir.

La opción de salir lleva al modo de Búsqueda.

Al seleccionar una opción el sistema direcciona al modulo correspondiente. Las instrucciones para operar sobre cada uno se encuentra en cada una de las siguientes secciones

ADMINISTRAR REPOSITORIOS

Cuando se selecciona Configurar del menú principal el sistema permite seleccionar un repositorio para indexar o buscar sobre él. También permite crear o eliminar un repositorio nuevo.

Seleccionar un repositorio

Para seleccionar un repositorio simplemente se debe hacer click sobre el nombre:

El repositorio seleccionado aparecerá en la parte superior derecha.

Crear repositorio

Para crear un repositorio nuevo se debe ingresar el nombre en el campo de texto que se encuentra debajo de todos los repositorios creados y oprimir el botón Adicionar.

Luego el sistema muestra el nuevo repositorio en la lista de repositorios creados.

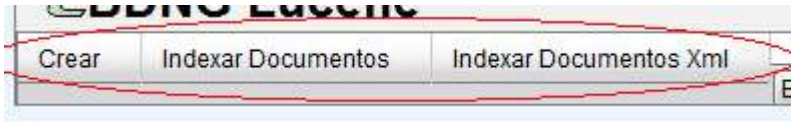
Eliminar repositorio

Para eliminar un repositorio simplemente se debe dar click en la opción eliminar como se muestra en la siguiente imagen:

El sistema solicita confirmar la eliminación del repositorio. Si el usuario confirma, se elimina el repositorio, de lo contrario vuelve a la página de Configuración.

INDEXACIÓN

Para este proceso se cuenta con el menú principal que se encuentra en la parte superior derecha con las siguientes opciones.

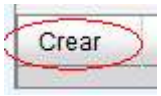


Al seleccionar una opción el sistema direcciona al modulo correspondiente. Las instrucciones para operar sobre cada uno se encuentra en cada una de las siguientes secciones

Crear un documento

Accediendo a esta opción se puede asociar un archivo con sus respectivos metadatos e indexar el contenido de dicho archivo.

Para ingresar a esta opción se debe ingresar por el menú principal.



El sistema presentará una lista con las opciones de los metadatos a ingresar y un campo en el cual el usuario deberá ingresar la información.

El usuario podrá asociar tantos metadatos como desee. Para agregar otro metadato se debe dar click en el botón “Añadir fila”.

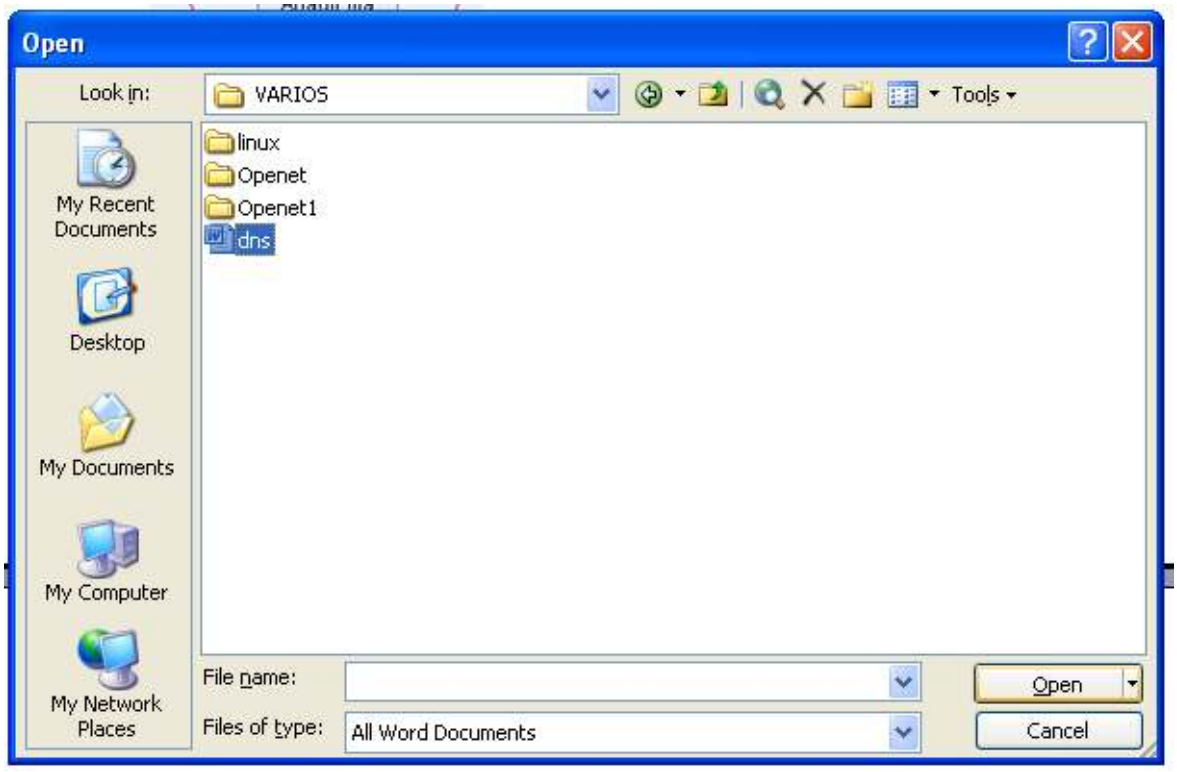
El sistema añadirá un registro para ingresar información.

Luego deberá seleccionar de la lista el metadato a asociar.

Y luego ingresar los datos correspondientes a dicho metadato y oprimir el botón “Save”.

También se cuenta con la opción de agregar un archivo digital. Para ello se debe oprimir el botón Browse.

Luego se deberá seleccionar el archivo según la ubicación en la estación de trabajo



Cuando se encuentren todos los datos listos se oprime el botón Guardar para almacenar el documento.

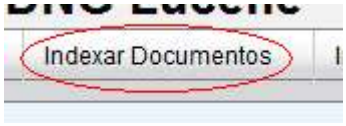
Nota: El archivo se almacenará en el servidor en la ruta configurada en el archivo *configuration.properties* en la propiedad *texto_dir*. Para esto se debe tener acceso al código fuente del sistema. Para más detalles, ver el documento Manual Técnico.

El sistema muestra si el documento fue indexado o no.

Indexar documentos

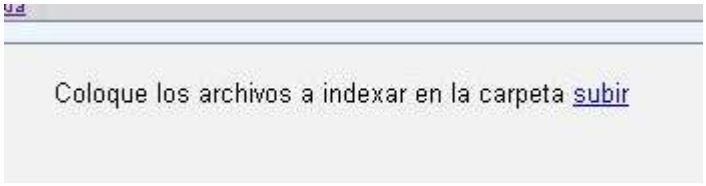
Al ingresar a esta opción se indexarán todos los documentos del directorio configurado en el archivo *configuration.properties* en la propiedad *texto_dir*.

Para ingresar a esta opción se debe ingresar por el menú principal.



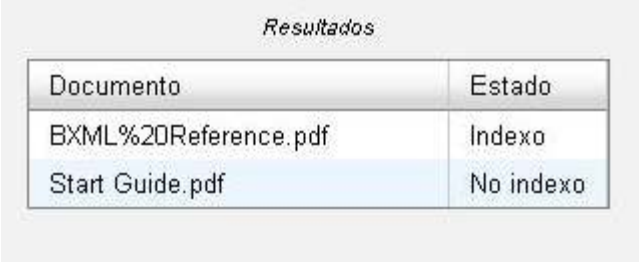
El sistema pedirá confirmar la acción. Si el usuario confirma entonces se indexarán todos los archivos contenidos en dicha carpeta y sus subcarpetas.

Se debe seleccionar el link “subir” para indexar todo el directorio.



Los archivos que indexa son los de extensión .doc, .pdf. .xls, .pps, .ppt, .html, .htm, .xhtml, .xml y .txt.

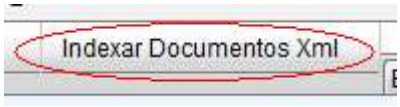
Luego de indexados todos los archivos, el sistema mostrará un listado de los documentos indexados y no indexados.



Indexar documentos XML Dublin Core

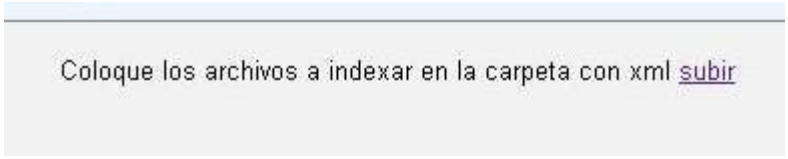
Al ingresar a esta opción se indexarán todos los documentos xml que contienen información de documentos referenciados por medio del modelo Dublin Core. El directorio a indexar será el directorio configurado en el archivo *configuration.properties* en la propiedad *texto_xml*.

Para ingresar a esta opción se debe ingresar por el menú principal.

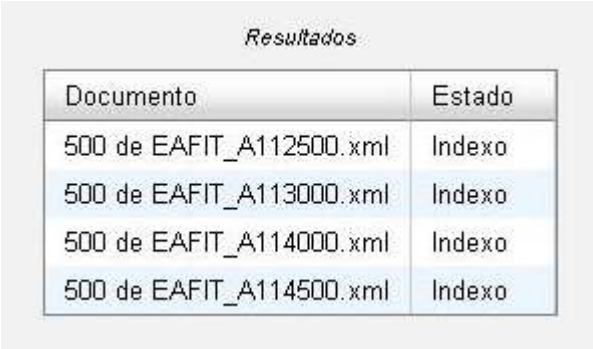


El sistema pedirá confirmar la acción. Si el usuario confirma entonces se indexarán todos los archivos contenidos en dicha carpeta y sus subcarpetas.

Se debe seleccionar el link “Subir” para indexar todo el directorio.

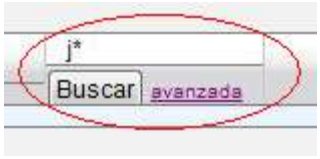


Luego de indexados todos los archivos, el sistema mostrará un listado de los documentos indexados y no indexados.



BUSQUEDA

Esta opción se encuentra en el menú principal. La búsqueda básica se puede realizar directamente desde dicho menú, o si se desea hacer una búsqueda avanzada se debe seleccionar dicha opción.

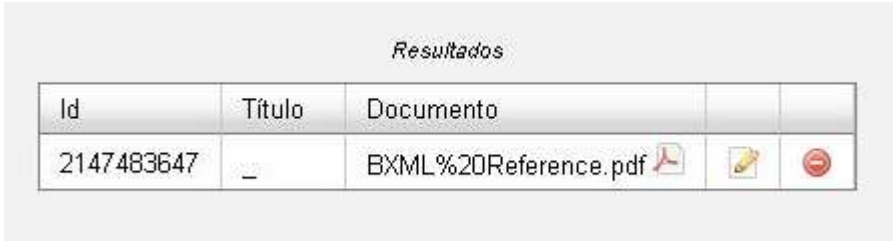


Búsqueda básica

Esta búsqueda consiste en ingresar los parámetros de búsqueda y el sistema se encargará de retornar los documentos coincidentes sin importar en que metadato o sección del documento se encuentran las palabras a buscar.

Esta búsqueda se realiza desde el menú principal, se debe ingresar la(s) palabra(s) a buscar y oprimir el botón buscar.

El sistema retornará los resultados que coincidan con esta búsqueda y los campos de Id, Titulo y Documento.

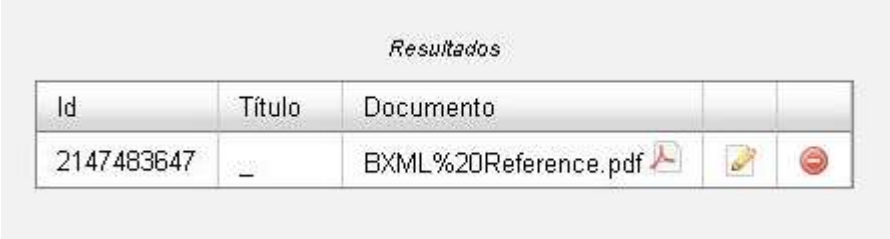


Búsqueda avanzada

Para ingresar a esta búsqueda se debe ingresar desde el menú principal.

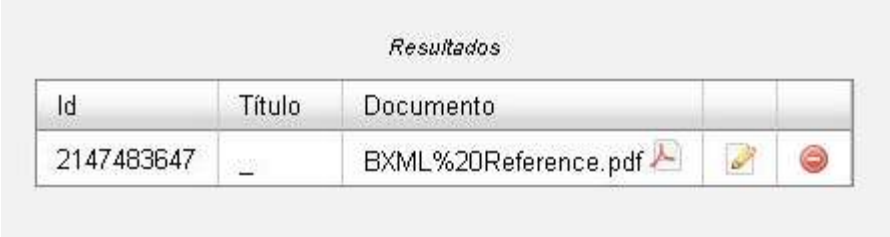
Para esta búsqueda se pueden seleccionar los campos específicos y las palabras que se desean buscar, y los operadores lógicos entre dichos campos para hacer más precisa la búsqueda.

Al oprimir el botón buscar se muestran los resultados según los criterios de búsqueda.

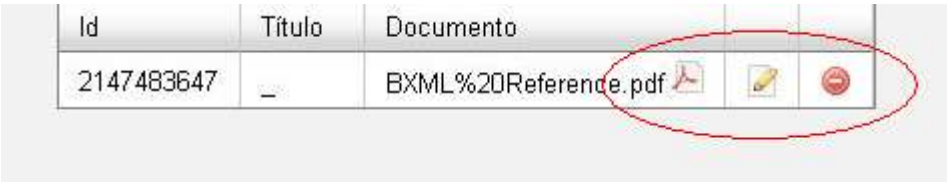


Ver resultados

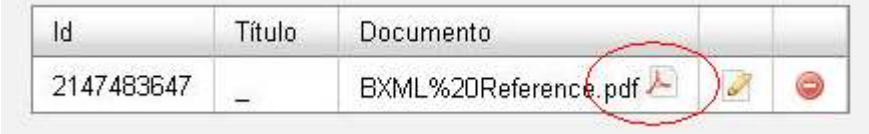
Sin importar el tipo de búsqueda, los resultados solo mostrarán algunos de los datos básicos asociados al documento.



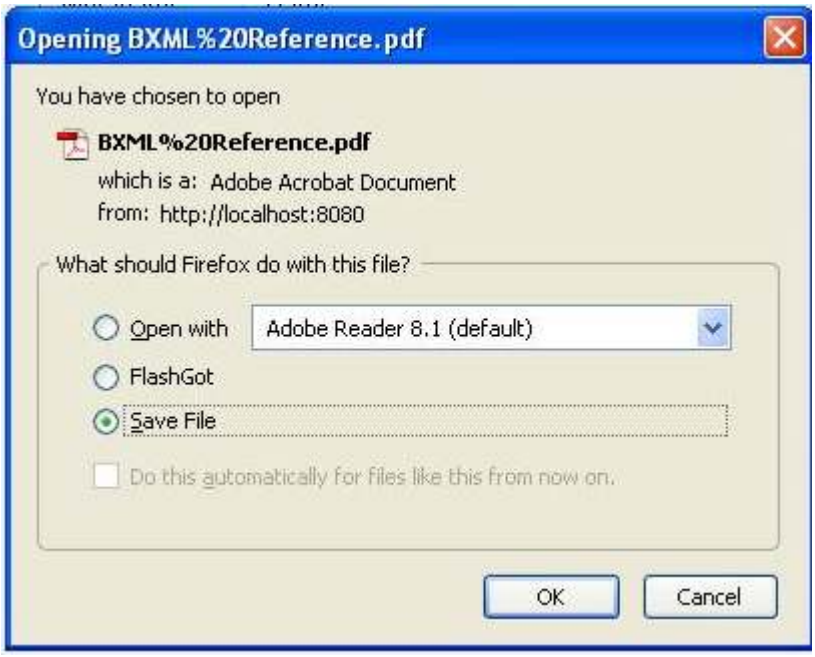
Cada registro encontrado tendrá la posibilidad de ver el detalle, descargar documento y eliminar.



Para descargar el documento se debe dar click sobre el icono del documento.

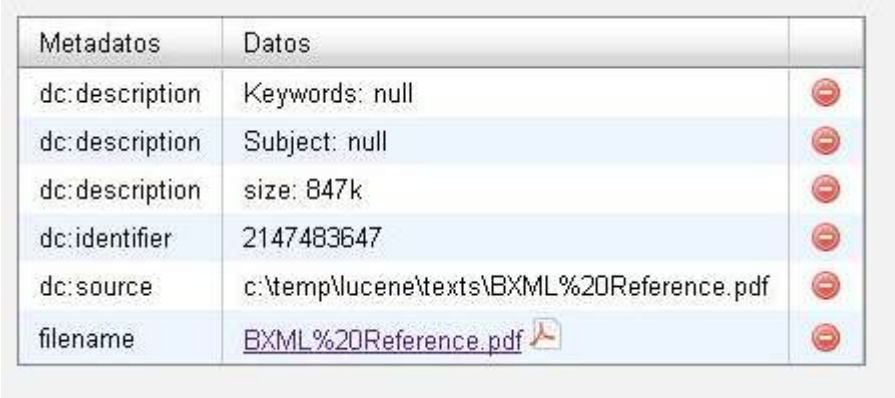


El sistema solicitará al usuario donde desea descargar el documento en la estación local, y allí será almacenado.



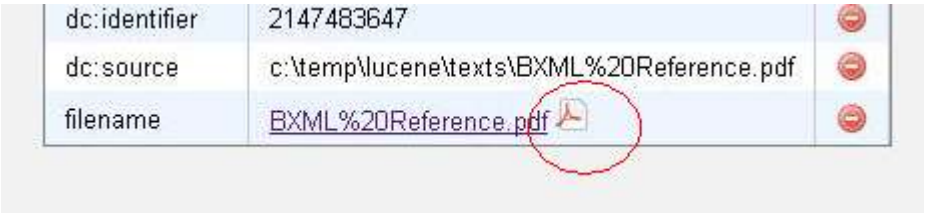
Ver detalle

En esta sección se muestran todos los datos que tiene el documento, si un dato se encuentra vacio entonces no se mostrará en el detalle.



Metadatos	Datos
dc:description	Keywords: null
dc:description	Subject: null
dc:description	size: 847k
dc:identifier	2147483647
dc:source	c:\temp\lucene\texts\BXML%20Reference.pdf
filename	BXML%20Reference.pdf

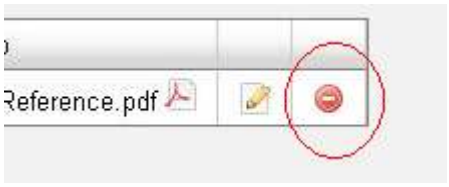
Para descargar el documento se debe dar click sobre el icono del documento.



Eliminar documento

Cuando se desea eliminar un documento se debe seleccionar la opción eliminar desde ver detalle o desde los resultados de la búsqueda:

Desde los resultados de búsqueda se debe oprimir el botón eliminar:



Desde ver detalle se debe oprimir el botón eliminar:

El sistema pide confirmación. Si el usuario confirma se elimina el documento y el archivo asociado si existe.